基于改进 YOLOv2 的无标定 3D 机械臂自主抓取方法 *

余玉琴 a，魏国亮 b，王永雄 a

(上海理工大学 a.光电信息与计算机工程学院; b.理学院，上海 200093;)

摘　要：提出了一种多物体环境下基于改进 YOLOv2 的无标定 3D 机械臂自主抓取方法。首先为了降低深度学习算法 YOLOv2 检测多物体边界框重合率和 3D 距离计算误差，提出了一种 YOLOv2 改进的算法。利用此算法对图像中的目标物体进行检测识别，得到目标物体在 RGB 图像中的位置信息；然后根据深度图像信息使用 K-means++聚类算法快速计算目标物体到摄像机的距离，估计目标物体大小和姿态，同时检测机械手的位置信息，计算机械手到目标物体的距离；最后根据目标物体的大小、姿态和到机械手的距离，使用 PID 算法控制机械手抓取物体。提出的改进 YOLOv2 算法获得了更精准的物体边界框，边框交集更小，提高了目标物体距离检测和大小、姿态估计的准确率。为了避免了繁杂的标定，提出无标定抓取方法，代替了基于雅克比矩阵的无标定估计方法，通用性好。实验验证了提出的系统框架能对图像中物体进行较为准确的自动分类和定位，利用 Universal Robot 3 机械臂能够对任意摆放的物体进行较为准确的抓取。

关键词：改进 YOLOv2；无标定；PID 控制；机械臂抓取

中图分类号：TP391.4　　doi: 10.19734/j.issn.1001-3695.2018.10.0821

3D uncalibrated robotic grasping method based on improved YOLOv2

Yu Yuqin[a], Wei Guoliang[b], Wang Yongxiong[a]

(a. *School of Optical-Electrical & Computer Engineering, b. College of Science, University of Shanghai for Science & Technology, Shanghai* 200093, *China*)

Abstract: This paper proposed an uncalibrated 3D robotic arm grabbing method based on improved YOLOv2 in a multi-object environment. Firstly, in order to reduce the depth learning algorithm YOLOv2 detection multi-object bounding box overlapping rate and 3D distance calculation error. It proposed an improved algorithm for YOLOv2. Using this algorithm to detect and identify the target object in the image, obtain the position information of the target object in the RGB image, and then use the k-means++ clustering algorithm to quickly calculate the distance from the target object to the camera according to the depth image information, and estimate the target object size and pose. Simultaneously, use the improved YOLOv2 to get the bounding box of the gripper and calculate the distance from the robot to the target object. Then the system estimates the distance between the fixture, camera and object in the manipulator coordinate system. Finally, the system uses the PID algorithm to control the gripper to grab the object according to the size and posture of the object and the distance from the object to the gripper. In this paper, the detected boundary boxes of the target object is more accurate based on the improved YOLOv2 than on old one. It also enhances the distance from the fixture to the object and the size of the object as well as the accuracy of the pose estimation. In addition, in order to avoid complicated calibration, this paper proposes a non-calibration method. This learning scheme is different from the traditional uncalibrated estimation method based on Jacobian matrix, because it has good universality. A simulation experiment shows that the proposed method can accurately classify and locate the objects in the image, The Universal Robot 3 robotic arm uses this framework to verify the effectiveness of capturing objects in a cluttered environment.

Key words: improved YOLOv2; uncalibration; PID control algorithm; robotic grasping

0　引言

基于视觉的智能机械臂物体抓取具有广泛的应用场景和较高的应用价值，物品分拣、垃圾分拣就是其典型任务。 传统的垃圾分拣工作采用人工分拣的形式，有些电子垃圾、化学品垃圾对人体危害较大。基于视觉的智能机械臂系统可自动识别不同种类、不同大小的垃圾，分辨出可再利用部分并自动实现分拣。 因此可用于快递分拣、工厂流水线上的零件摆放，逐步代替人工从事劳动强度大、重复单调的工作，同样也是工业 4.0 和人工智能的主要研究方向之一。

在传统的机械臂物体抓取中，对位姿固定的单目标物体采用人工示教的方式抓取。在常规的视觉伺服中， 摄像机和机械臂的位置相对固定，目标物体单一且位姿固定；由于不能自主感知工作环境、物体类别、形状、尺寸和位姿等信息，传统机械臂系统的物体抓取方法具有诸多不确定性，只能在特定环境下使用。多个物体存在、物体的种类不同、大小不同、物体的位姿变化、摄像机和机械臂的相对位置不固定等问题使得传统视觉机械臂系统无法完成复杂的抓取任务。

为了能够在自然环境中实现自主物体抓取，研究人员不断改进基于视觉的机械臂物体抓取方法。文献[1]介绍了传统

收稿日期：2018-10-22；**修回日期：**2019-01-07　　**基金项目：**国家自然科学基金资助项目（61673276）；上海市科委地方能力建设项目（15550502500）

作者简介：余玉琴（1994-），女，安徽霍山人，硕士研究生，主要研究方向为智能机器人与视觉(yuqinyus@163.com)；魏国亮（1973-），男，教授，博士，主要研究方向为复杂系统协同控制；王永雄（1970-），男，副教授，博士，主要研究方向为智能机器人与视觉、模式识别.

的基于图像二值图位置检测的机械臂物体抓取方法，文献[2~4]提出了一种基于雅可比矩阵估计的视觉伺服控制方案，以上方法都是单个物体场景下，机器人和摄像机的位置相对固定。在视觉识别阶段，大多数方法还是手工设定特征，但是在实际的抓取任务中，目标物体的大小、形状、外部光照强度、角度变化和采样角度不确定，传统的特征提取方法提取的物体特征鲁棒性差，不能适应新物体和多变的环境。

2012 年，Hinton 课题组使用深度学习方法 AlexNet[5]，在 ImageNet 图像识别比赛中夺得冠军，此后深度学习迅速受到广泛的关注，并逐步应用于机械臂物体抓取领域。文献[6]提出在抓取姿态不确定的情况下，使用卷积神经网络学习抓取函数，此方法泛化能力强、能够适应新物体，但都只适用于单个物体场景，无法在多个物体共存的杂乱环境下应用。因此，研究人员提出了多个物体杂乱共存环境下的物体抓取方法。文献[7]提出基于深度学习的多视图、自监督方法来估计物体 6D 姿态，完成物体抓取。此方法需要通过繁杂的视觉标定确定摄像机和机械臂之间的相对位置。 针对此问题，文献[8]提出了一种基于深度学习无标定手眼协调抓取方法。他们使用了 6 到 14 个机器人，经历 3 个月收集了超过 80 万次的抓握尝试数据集，训练了一个深层卷积网络控制机械臂抓取物体。 此方法泛化能力强，但此方法的数据集收集难度大，成本高，并且此方法只适用于某一特定型号的机械臂。

针对多目标、环境杂乱、物体位姿不固定、大小不固定、摄像机和机械臂相对位置不固定的抓取环境，本文提出了一种 YOLOv2 改进的算法，实现杂乱环境中多物体的自动检测，克服原算法检测物体框重叠率过高等问题，并识别物体的类别，估计目标物体的边界框信息、大小和姿态，同时检测机械手的边界框，然后使用 K-means++聚类算法快速计算摄像机、机械手和物体三者之间的距离，最后根据目标物体的大小、姿态及机械手到目标物体的距离，使用 PID 控制算法控制机械手抓取物体。本方法的创新性如下：

a)提出了一个全新的基于机器视觉和机器学习的无标定 3D 机械臂抓取框架，首先使用深度学习检测物体，获取物体大致位姿，再采用 K-means++聚类算法计算摄像机、机械臂和物体三者之间的距离，最后利用 PID 控制方法实现物体抓取。此方法通用性好。使用机器学习方法获得位姿信息，代替了传统的无标定视觉伺服中的雅克比矩阵估计，优点是计算简便，实时性好。

b）改进了 YOLOv2 算法的物体边界框确定方法，改进后检测出的目标物体边界框与邻近物体边界框的交集更小，提高了目标物体到摄像机之间距离的计算精度，进而提高了物体的姿态估计精度和大小估计精度。

c）相比于传统的人工特征提取方法，采用改进 YOLOv2 方法学习目标物体的特征，借助深度学习预训练，能够适应没有经过训练的新物体，具有较高的泛化能力和稳定性。 相比于传统的抓取位置检测的方法，利用图像信息和深度信息进行无标定 PID 控制，避免了摄像机和机械臂基座之间相对位置的繁复标定。相对于大规模数据集的手眼协调无标定抓取方法[8]，避免使用成本高昂的设备收集数据集，此方法和人类抓取物体的方法更相似，实现过程更经济、更快捷， 符合人工智能自主抓取的理念。

1　系统框架和流程

系统框架及算法流程如图 1 所示。该系统主要包括 RBG-D 摄像机（Kinect2.0）和 UR3 机械臂，摄像机固定在工作台一端，摄像机成像平面垂直于工作台桌面。此抓取方法主要包括三部分：物体检测算法、目标物体姿态和大小估计、物体抓取控制。首先使用摄像机采集目标物体的彩色图像和深度图像，将 RGB 图像输入到改进 YOLOv2 物体检测算法中，检测并识别系统空间下各个物体的类别、目标物体在 RGB 图像中的位置和边界框； 然后结合深度图像使用 K-means++聚类算法快速计算目标物体到摄像机的距离， 根据目标物体的距离和边界框估计目标物体的姿态和大小；通过使用聚类算法实时计算机械手到摄像机的距离，获得目标物体和机械手在机械臂坐标系 XR 方向上的距离。依据 RGB 图像计算目标物体和机械手在机械臂坐标系 YR 方向上的距离。最后根据目标物体的大小和姿态调节机械手， 采用以 XR 和 YR 方向上的距离作为输入的 PID 闭环控制算法实现机械臂抓取物体。

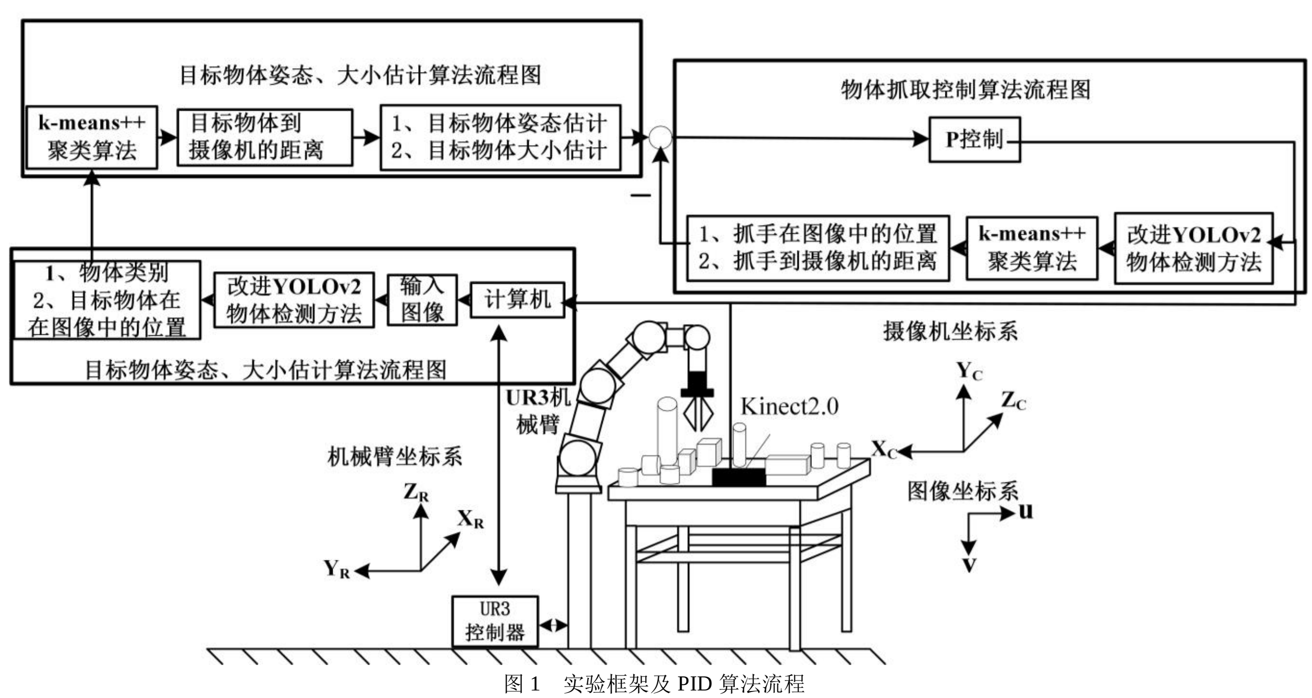


图 1　实验框架及 PID 算法流程

Fig. 1　Experimental framework and algorithm flow

2　目标检测算法

经典的目标检测算法(R-CNN)[9]首先用选择性搜索方法[10]在图像中搜素 1k~2k 个候选框，再使用卷积神经网络(CNN)模型提取特征。每一个候选框都需要输入到 CNN 模型中提取特征，上千个候选框存在大量的范围重叠，重复的特征提取产生巨大的计算量， 使得目标检测不具备实时性。Faster-RCNN[11]实现了较快速的目标检测，此方法提高了目标检测的精度和速度，但是候选框的生成和分类过程计算量大，无法达到实时检测目标。

文献[12]提出了 YOLO 物体检测方法，此算法将物体检测任务当做一个回归问题来处理，将整张图片输入到 YOLO 网络，此方法的优点是检测物体速度很快，能够有效的避免背景错误、学习物体的泛化特征，但其物体检测精度低，对于密集的小物体检测效果差。为了让目标检测算法同时具备检测速度快和检测精度高的优点，文献[13]提出了 YOLOv2 物体检测算法。使用 VOC 数据集训练 YOLOv2 模型，mAP(mean average precision)为 76.8，检测速度为 67FPS。VOC 数据集训练 Faster R-CNN 模型， mAP 为 73.2[13]，检测速度为 7FPS。YOLOv2 的检测精度优于 Faster R-CNN，检测速度快于 YOLO。所以本文最终采用 YOLOv2 作为机械臂抓取任务中的目标检测模型。

2.1　YOLOv2 目标检测模型

YOLOv2 的分类网络是 Darknet－19 网络模型，由 19 个卷积层和 5 个池化层组成，大多使用 3×3 的滤波器，每个池化操作后使通道数加倍，使用全局平均池化做预测[14]，使用 1×1 滤波器来压缩卷积之间的特征表示[15]。使用批归一化来稳定训练，加速收敛，并正则化模型[16]。YOLOv2 的检测网络使用了 Anchor 预测框的卷积层，并且使用 k-means 聚类算法优化了先验预测框的选取，去掉了全连接层，使得 YOLOv2 能够准确快速的检测物体。

YOLOv2 采用多任务损失来最小化目标函数， 目标函数定义为

$$\begin{aligned}&\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2]\\&+\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}\left[(\sqrt{w_i}-\sqrt{\hat{w}_i})^2+(\sqrt{h_i}-\sqrt{\hat{h}_i})^2\right]\\&+\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}(C_i-\hat{C}_i)^2+\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}(C_i-\hat{C}_i)^2\\&+\sum_{i=0}^{S^2}1_i^{obj}\sum_{c\in classes}(p_i(c)-\hat{p}_i(c))^2\end{aligned} \tag{1}$$

其中: i 为单元网格的索引， B 表示一个网格单元预测的边界框数目，此处 B=5， j 为网格单元预测的边界框索引。1_i^{obj} 表示目标是否出现在网格单元 i 中，1_{ij}^{obj} 表示网格单元 i 中的第 j 个边界框预测器负责该网格单元的目标预测，1_{ij}^{noobj} 表示网格单元 i 中的第 j 个边界框预测器不负责该网格单元的目标预测。(x,y) 为中心点的归一化偏移坐标， w,h 分别为边界框归一化的宽度和高度， $(\hat{x},\hat{y})$ 为图片中物体真实边界框的中心点， $\hat{w},\hat{h}$ 为物体真实边界框归一化的宽度和高度， C 为预测的单元格的置信率， $\hat{C}$ 为真实的单元格的置信率， p 为预测的物体置信率， $\hat{p}$ 为真实的物体置信率。 $classes$ 为待检测物体的类别数目， $\lambda_{coord}=5$， $\lambda_{noobj}=0.5$ 。

2.2　改进的 YOLOv2 目标检测模型

YOLOv2 物体检测算法输出待测物体边界框的中心点在图像的位置 (b_x,b_y) 、边界框的宽 t_w 与高 t_h 、置信率 p 。在自制数据集中，需要对每张图片中的 m 个进行标注，标注内容包括各类物体的类别 $Class_i$ 、各类物体的边界框的长 t_{w_i} 和宽 t_{h_i} 及边界框的中心点坐标 (b_{x_i},b_{y_i}) ， $i=1\cdots m$ ，边界框内仅包含物体。在复杂环境下检测物体，物体与物体之间距离太近时，边界框有重合部分。重合率过高时，采用深度信息计算物体到摄像机的距离会产生较大的误差。为了解决此问题，提出了一种改进的 YOLOv2 检测模型，将原 YOLOv2 输出的边界框的宽与高分别缩小 k_w、k_h 倍， k_w、k_h 计算方法如下：

a) 标注训练集中多张图片 N 个物体的边界框 $t_{w_train_i},t_{h_train_i}$ $i=1\cdots N$ 。

b）使用原 YOLOv2 方法检测训练集 N 物体的边界框的 t_{w_i},t_{h_i} $i=1\cdots N$ 。

c) 使用以下公式计算每个物体的边界框对应的 k_{w_i},k_{h_i} $i=1\cdots N$ 。

$$\begin{cases}k_{w_i}=t_{w_i}/t_{w_train_i}\\k_{h_i}=t_{h_i}/t_{h_train_i}\end{cases} \tag{2}$$

d)计算

$$\begin{cases}k_w=\sum_{1}^{N}k_{w_i}\Big/N\\k_h=\sum_{1}^{N}k_{h_i}\Big/N\end{cases} \tag{3}$$

改进的 YOLOv2 模型如图 2 所示。使用改进的 YOLOv2 方法大幅度减小了物体间边界框的重合率。改进后的模型更加适用于复杂多目标下的目标检测。改进后的效果示意图如图 3 所示。

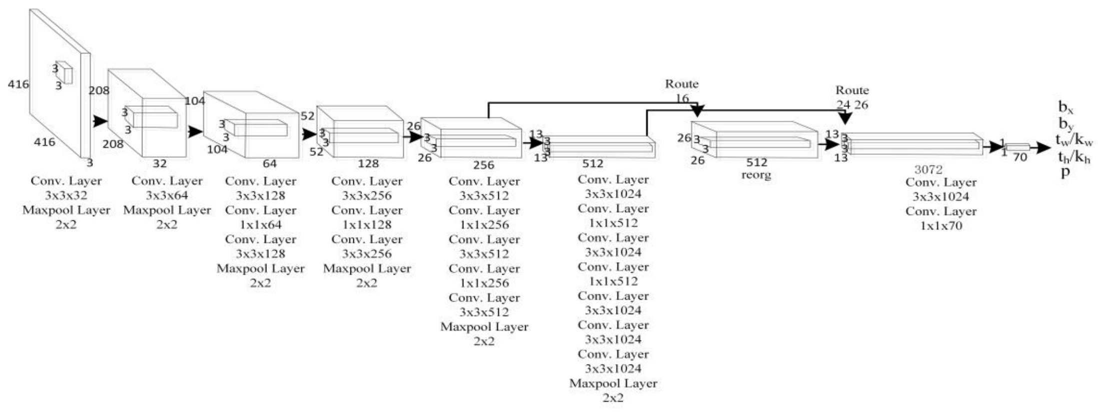


图 2　改进的 YOLOv2 网络结构

Fig. 2　Improved yolov2 network structur

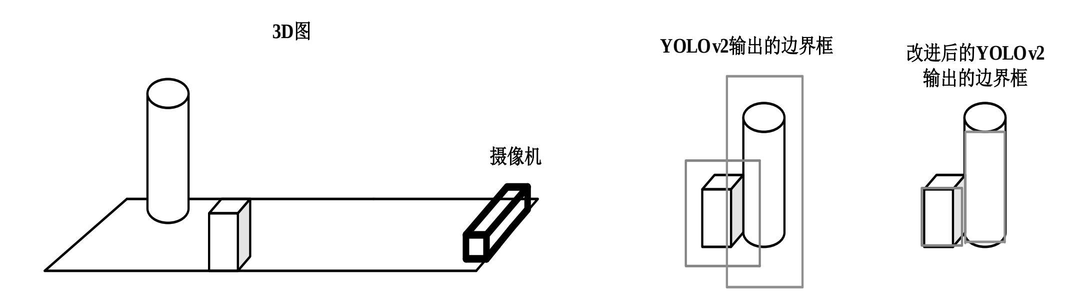


图 3　改进 YOLOv2 模型检测效果示意图

Fig. 3　Improved yolov2 model detection effect

3　物体距离和物体大小、姿态估计

3.1　基于 K-means++的物体距离计算

经典的 K-means 聚类算法具有聚类效果不佳和收敛速度慢等问题，难以保证机械臂抓取物体的实时性。本文选用 K-means++聚类算法[18]。它采用初始中心点彼此尽可能远离的策略来解决上述问题，将 n 个样本点聚类为 k 类算法如下：

初始化一个空的集合 $\boldsymbol{M}$ ，用于存储选定的中心点。

a)从输入样本中随机选定第一个中心点 $\mu^{(j)}, j\in\{1,\cdots k\}$ ，并将其加入到集合 $\boldsymbol{M}$ 中。

b)对于集合 $\boldsymbol{M}$ 之外的任一样本点 $x^{(i)}, i\in\{1,\cdots n\}$ ， 通过计算找到与其平方距离最小的样本 $d(x^{(i)},\boldsymbol{M})^2$ 其中：

$$d(x^{(i)},\boldsymbol{M})^2=(x^{(i)}-\boldsymbol{M})^2=\left\|x^{(i)}-\boldsymbol{M}\right\|^2 \tag{4}$$

c)计算每个样本点成为下一个聚类中心的概率：

$$P_i=\frac{d(x^{(i)},\boldsymbol{M})^2}{\sum_{i=1}^{n}d(x^{(i)},\boldsymbol{M})^2} \tag{5}$$

按照轮盘法选择出下一个聚类中心点 $u^{(p)}$ ， 并将其加入到集合 $\boldsymbol{M}$ 中。

d)重复步骤 b)c)，直到选定 k 个中心点。

e)将每个样本点划分到距离它最近的中心点 $u^{(j)}$ 所代表的簇中。

f)将各簇中所有样本点的中心代替原来的中心点。

g)重复步骤 e)f)使得簇内误差平方和 SSE 最小，直到中心点不变或者达到预期迭代次数时，算法中止。

$$SSE=\sum_{i=1}^{n}\sum_{j=1}^{k}w^{(i,j)}\left\|x^{(i)}-\mu^{(j)}\right\|_2^2 \tag{6}$$

如果样本 $x^{(i)}$ 属于簇 j ， 则 $w^{(i,j)}=1$ ， 否则 $w^{(i,j)}=0$ 。

本文首先利用改进的 YOLOv2 得到目标物体的边界框，根据边界框内每个像素对应的深度值，基于上述 K-means++步骤，将深度值快速聚类为三类， 再将三个聚类中心值按照升序排列，选择排序第 2 的聚类中心值作为物体到摄像机的距离。

3.2　物体的大小、姿态估计

3.2.1 物体大小估计

根据改进 YOLOv2 检测的目标物体边界框的长宽 b_{object_w} 、 b_{object_h} ，取其中的较小值作为目标物体在图像。坐标系中的宽度值为

$$W_{i_object}=\begin{cases}b_{object_w}, & b_{object_w}<b_{object_h}\\ b_{object_h}, & b_{object_h}\le b_{object_w}\end{cases} \tag{7}$$

物体大小估计模型如图 4 所示，根据相似三角形性质可以列出如下等式：

$$\frac{W_{i_object}}{f}=\frac{W_{r_object}}{d_{object}} \tag{8}$$

根据等式可以计算出目标物体真实的宽度为

$$W_{r_object}=\frac{W_{i_object}f}{d_{object}} \tag{9}$$

其中： f 为摄像机的焦距。

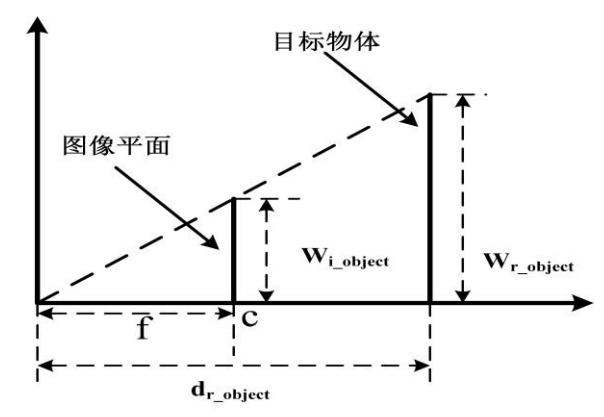


图 4　物体大小估计模型

Fig. 4　Object size estimation model

3.2.2 物体姿态估计

基于机械臂有三种抓取姿态，本文将目标物体的姿态也对应分为三种。根据改进 YOLOv2 检测的目标物体的边界框的长和宽分别为 b_{object_w} 和 b_{object_h} ，计算物体的长宽比 r_{object} 来估计目标物体的姿态。

$$r_{object}=b_{object_h}/b_{object_w} \tag{10}$$

将 r_{object} 分为三段，进而估计出目标物体的姿态：

$$目标物体姿态=\begin{cases}A & r_{object}\le r_1\\ B & r_1<r_{object}\le r_2\\ C & r_{object}>r_2\end{cases} \tag{11}$$

其中： r_1、 r_2 为分段系数，根据实验经验可得。三种抓取姿态如图 5 所示。

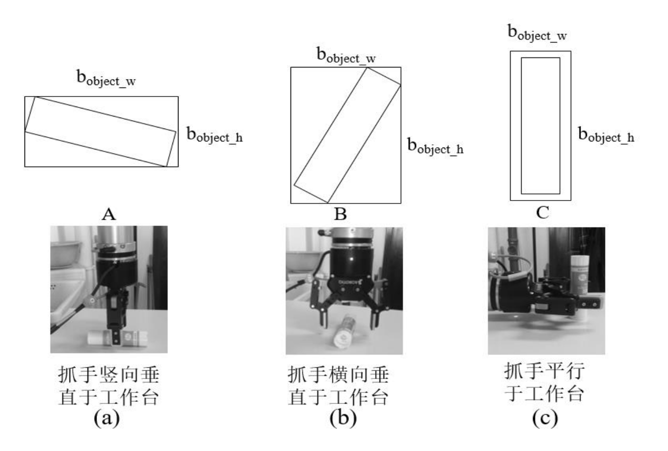


图 5　物体的三种姿态及对应的机械手抓取状态

Fig. 5　Three poses of the object and the corresponding grabbing state

4　无标定闭环抓取控制

为了实现快速无标定的视觉物体抓取，本文提出采用 PID 控制方法实现无标定闭环抓取，主要包括以下三步：

a）通过改进的 YOLOv2 算法检测目标物体的边界框 b_{object_x}、b_{object_y}、b_{object_w}、b_{object_h} 及目标物体的类别 $Class_{object}$。使用 K-means++聚类算法计算出目标物体到摄像机的距离 d_{object}。

b）根据长宽比 r_{object} 判断目标物体的姿态，根据姿态调整机械手的抓取姿态。实时检测机械手的边界框 $b_{robotiq_x}$、$b_{robotiq_y}$、$b_{robotiq_w}$、$b_{robotiq_h}$，通过 K-means++聚类算法计算出机械手到摄像机的距离 $d_{robotiq}$，使用 PID 控制算法控制机械臂移动，不断靠近目标物体，直到误差小于给定阈值。

$$\begin{cases}|b_{object_x}-b_{robotiq_x}|\le Thresh_{image}\\ |d_{object}-d_{robotiq}|\le Thresh_{distance}\end{cases}\tag{12}$$

其中：$Thresh_{image}$ 为图像坐标系下目标物体和机械手的中心点的误差阈值，$Thresh_{distance}$ 为摄像机坐标系下目标物体和机械手的距离差阈值。

c）控制机械臂垂直向下移动到距离桌面 *1*cm 处，根据估计的物体宽度 W_{r_object} 控制机械手闭合，抓取物体并移动到存放物体位置，打开机械手，完成抓取。

为了使机械臂末端到达目标点的运动时间最短，本文设置机械臂的运动速度为机械臂能够承受的最大速度，并且设置机械臂的运动路径为直线路径，即机械臂末端以最大速度沿直线从当前点运动到目标点。此方法减少了机械臂末端到达目标点所需时间，提高了机械臂无标定闭环抓取的效率。

本文使用 PID 控制算法控制机械臂移动，计算期望机械手位置与当前机械手位置差值，利用差值使用 PID 控制算法控制机械手靠近目标物体，再次计算期望机械手位置与反馈回来的当前机械手位置差值，使用 PID 控制算法控制机械手移动，直到机械手运动到目标物体正上方位置。机械手位置图如图 6 所示。

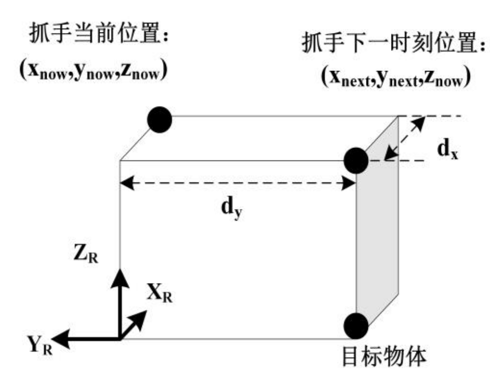


图 6　机械手位置图

Fig. 6　Location of gripper

机械手在机械臂坐标系需要移动距离的数学表达式如下：

$$\begin{cases}d_x=(d_{object}-d_{robotiq})p_x\\ d_y=(b_{object_x}-b_{robotiq_x})p_y\end{cases}\tag{13}$$

其中：d_x 表示机械手在机械臂坐标系 X_R 方向上需要移动的距离，d_y 表示机械手在机械臂坐标系 Y_R 方向上需要移动的距离。p_x, p_y 为 PID 控制算法系数，p_x 和目标物体距离摄像机的距离成正比，即

$$p_x=d_{object}p'_x\tag{14}$$

p'_x 和 p_y 根据实验经验可得。机械手的目标位置表达式如下：

$$\begin{cases}x_{next}=x_{now}+d_x\\ y_{next}=y_{now}+d_y\end{cases}\tag{15}$$

其中：x_{next} 为机械手下一时刻的在机械臂坐标系 X_R 方向上的位置，y_{next} 为机械手下一时刻的在机械臂坐标系 Y_R 方向上的位置，x_{now} 为机械手当前时刻在机械臂坐标系 X_R 方向上的位置，y_{now} 为机械手当前时刻在机械臂坐标系 Y_R 方向上的位置。为了防止机械手触碰到目标物体，改变了目标物体位置，所以在机械手到达目标物体的正上方之前，不改变机械手在 Z_R 方向上的位置。

5　实验结果及分析

机械臂抓取仿真实验环境为 Ubuntu 16.04 系统，摄像机的图像采集环境为 ROS Kinetic，改进 YOLOv2 物体检测算法、K-means++聚类算法与 PID 控制算法的编程环境都是在标准 Python 环境 IDLE 集成开发环境中进行实验。

本实验按照 PASCAL VOC 数据集格式自建数据集，用于训练目标检测模型。数据集图片采集使用 Kinect2.0，仅使用摄像机采集的彩色图像。在目标类别检测阶段，使用的数据集包括九类物体，共 9 000 张图片，其中九个类别分别为“battery”“cream”“jar”“chutty”“lotions”“bag”“box”“sols”“robotiq”。用于模型训练的图片每类随机各选取 800 张，每类剩余 200 张用于模型测试。

5.1　仿真实验

5.1.1 目标检测实验

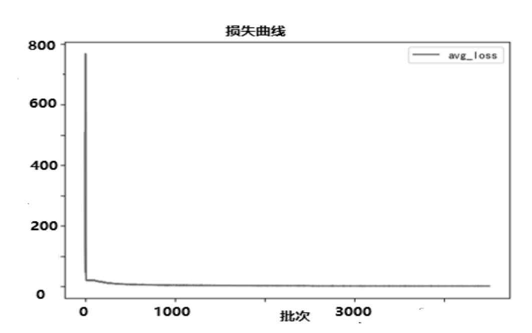
在对改进的 YOLOv2 目标检测网络进行训练时，使用显卡 GTX1080 对训练过程进行加速。改进的 YOLOv2 模型所依赖的深度学习框架为 Darknet，模型训练所需的参数设置如下：基础学习率为 0.001；最大迭代次数(max_batches)为 45000；学习率的衰减策略(policy)为“steps”，步长为“100，25000，35000”，相对于当前学习率的变化比率(scales)为“10，0.1，0.1”；每次迭代输入的图片数量(batch)为“64”，图片的子集数目 subdivision 为“8”；动量(momentum)为“0.9”；权重衰减率(decay)为“0.0005”；训练结果如图 7 所示。图 7 中，(a)中的损失曲线显示该模型的损失函数最终稳定趋于 0，(b)的区域平均 IOU 曲线显示平均 IOU 稳定在 0.7~0.85。从这两幅曲线图可以看出，整个模型的检测效果较好。

改进 YOLOv2 目标检测算法的检测速度约为 *0.014 592* s/张。目标检测实验结果如图 8 所示。目标检测实验准确率如表 1 所示。检测结果显示：在多物体共存环境下，未改进的 YOLOv2 检测得到的物体的边界框重合部分很大，改进后 YOLOv2 检测得到的边界框重合部分极小。

5.1.2 目标物体到摄像机距离计算

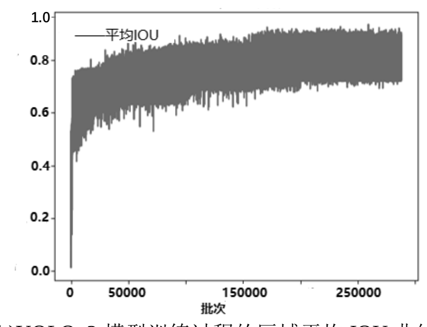
图 9 为使用 K-means++聚类算法对“sols”目标物体边界框对应的深度值进行聚类的结果图，第一类中心值为 *595* mm，第二类中心值为 *603* mm，第三类中心值为 *612* mm，选择第二类中心值 *603* mm 作为目标物体“sols”到摄像机的距离。

在多物体环境下，首先根据改进的 YOLOv2 检测识别出的图像中目标物体的类别、置信率、和边界框的中心坐标、边界框的长和宽。将预测的物体去除机械手，按照置信率大小降序排列，选取置信率最高的物体作为目标物体，根据目标物体的边界框信息，利用 K-means++聚类算法计算该物体到摄像机的距离，并将此距离和物体边界框信息保存下来。因为在抓取的过程中机械臂和机械手运动可能会遮挡住目标物体，所以在最开始需要将目标物体的边界框信息和距离信息保存下来，避免由遮挡引起的边界框误差和距离误差。



(a)YOLOv2 模型训练过程的损失曲线

(a)Loss curve during YOLOv2 model training



(b)YOLOv2 模型训练过程的区域平均 IOU 曲线

(b)Regional average IOU curve of YOLOv2 model training process

图 7 YOLOv2 训练过程损失值变化和区域平均 IOU 值变化曲线

Fig. 7 YOLOv2 training process loss value curve and regional average IOU value curve



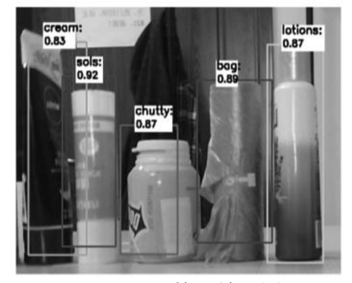
(a) YOLOv2 检测效果图 1

(a)Results of the YOLOv2 detection and identification 1



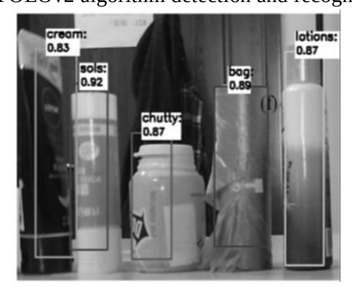
(b)改进后的 YOLOv2 检测效果图 1

(b)Improved YOLOv2 algorithm detection and recognition results 1



(c)YOLOv2 检测效果图 2

(c)Results of the YOLOv2 detection and identification 2



(d)改进后的 YOLOv2 检测效果图 2

(d)Improved YOLOv2 algorithm detection and recognition results 2

图 8 物体检测结果

Fig. 8 Object detection result

表 1 改进 YOLOv2 物体检测的准确率

Table 1 Accuracy based on improved YOLO2 detection algorithm

类别	battery	cream	jar	chutty	lotions	bag	box	sols	robotiq
准确率 (%)	81.9	80.2	87.5	94.0	95.1	95.7	96.1	95.54	93.86

为了验证改进 YOLOv2 能够有效提高目标物体到摄像机距离的计算精度，本文使用两类物体做了七组对比实验，物体摆放位置如图 8（a）所示，“cream” 在“lotions”后方距离恒定为 *40*mm，放置“lotions”到摄像机的距离为：*450*mm，*500* mm，*550* mm，*600* mm，*650* mm，*700* mm，*750*mm。每组实验的距离计算了五次，取五次的平均值作为最终的距离。实验结果如表 2 所示。根据表 2 计算得到，使用改进的 YOLOv2 输出的边界框计算的距离平均相对误差为 0.38%，平均距离绝对误差为 *2.3257* mm。

5.2 机械臂抓取实验

本实验不需要对摄像机和机械臂的相对位置进行繁杂的标定，不需要计算目标物体在机械臂坐标系下准确的 3 维位置信息，只需要确定摄像机坐标系和机械臂坐标系的关系，通过计算机械手和目标物体在图像中的位置、计算机械手和目标物体到摄像机的距离便可以完成抓取。并且摄像机在工作台上前后、左右移动适当的距离不影响抓取， 也不需要调整任何参数。

本实验环境模拟实际工业生产环境，实验台周围布置了很多的干扰物体，这些干扰物体颜色、形状、大小各异，用来测试本文提出方法在实际工业生产环境中进行物体分类抓取的鲁棒性。抓取实验采用 6 自由度人机协作型工业机械臂 UR3，机械臂实物图如图 10 所示，机械手采用 ROBOTIQ 二指夹手。安装在机械臂末端上。该机械臂可在半径 *600* mm 的可到达范围内任意移动。摄像机的测距范围为 *0.4 ~3* m，相机和目标物体的距离必须超过 *0.4* m，使得相机能够“看”到工作台的全部即可。

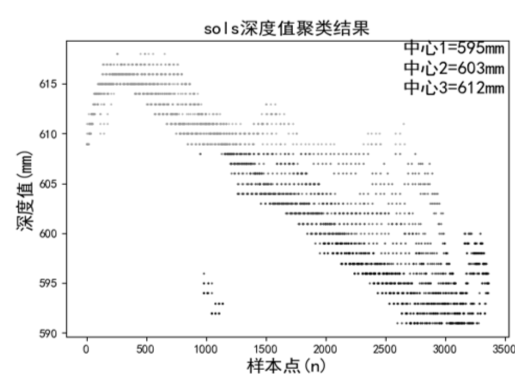


图 9 “sols”目标物体深度信息的聚类结果

Fig. 9 Clustering result of depth information of "sols" target object

表 2　距离计算的对比实验结果

Table 2　Comparison of experimental results of distance calculation

组别		1	2	3	4	5	6	7
真实距离值/mm	cream	490	540	590	640	690	740	790
	lotions	450	500	550	600	650	700	750
YOLOv2 得到的距离/mm	cream	490.57	510.45	557.19	612.67	657.06	739.59	755.90
	lotions	490.46	493.92	547.13	602.90	648.10	705.28	755. 16
改进 YOLOv2 得到的距离/mm	cream	490.46	539.78	591.28	643.14	688.34	749.90	792.62
	lotions	456.68	502.02	548.71	600.37	651.70	699.76	749.00
YOLOv2 得到的距离绝对误差/mm	cream	0.57	29.54	32.81	27.33	32.94	0.41	34.10
	lotions	40.46	6.08	2.87	2.90	1.90	5.28	5.16
改进 YOLOv2 得到的距离绝对误差/mm	cream	0.46	0.22	1.28	3.14	1.66	9.90	2.62
	lotions	6.68	2.02	1.29	0.37	1.70	0.24	1.00
YOLOv2 得到的距离相对误差/%	cream	0.12	5.47	5.56	4.27	4.77	0.06	4.32
	lotions	8.99	1.21	0.52	0.48	0.29	0.75	0.69
改进 YOLOv2 得到的距离相对误差/%	cream	0.09	0.04	0.22	0.49	0.26	1.33	0.33
	lotions	1.48	0.40	0.23	0.06	0.26	0.03	0.13

实验前，根据摄像机放置的位置，图像坐标系的 u 坐标轴的方向与机械臂坐标系 Y_R 坐标轴方向，摄像机坐标系 Z_C 坐标轴的方向与机械臂坐标系 X_R 坐标轴方向的对应的关系已确定，即当机械手和目标物体的边界框的中心点在图像坐标系的 u 轴方向上的距离满足以下条件时：

$$\left|b_{object_x}-b_{robotiq_x}\right|>Thresh_{image} \tag{16}$$

控制机械臂在机械臂坐标系下的 Y_R 坐标轴方向上朝着靠近目标物体的方向运动;当机械手和目标物体在摄像坐标系下 Z_C 坐标轴方向上的距离满足以下条件时：

$$\left|d_{object}-d_{robotiq}\right|>Thresh_{distance} \tag{17}$$

控制机械臂在机械臂坐标系下的 X_R 轴方向上朝着靠近目标物体的方向运动。

抓取步骤如第 4 章 a）~c）所述， UR3 机械臂抓取实验结果如图 10 所示。



(a)机械臂的初始位姿， 目标物体为电池

(a) Initial pose of the arm, the target object is the battery



(b)机械手移动到目标物体的正上方

(b) Gripper is directly above the target object



(c)机械手抓住目标物体

(c) Grab the target object



(d)为机械手抓起目标物体

(d) Target object is being grabbed by the manipulator

图 10　UR3 机械臂抓取实验结果

Fig. 10　Grasping experiment results of UR5 robot

6　结束语

本文采用改进的 YOLOv2 实现了在杂乱环境下对不同种类、不同尺寸的物体分类和定位，利用 K-means++聚类算法获得目标物体到摄像机距离，并提高了目标物体大小、姿态和目标物体到机械手距离的估计精度，最后使用无标定的 PID 控制方法实现抓取，避免了繁杂的标定。在多目标、环境杂乱、目标物体位姿、大小不固定、摄像机和机械臂相对位置不固定的抓取环境下，实验验证了改进的 YOLOv2 检测方法能够对目标物体实现较为准确分类和定位。未来的研究方向是优化本文方法，包括使用的数据集标注框优化、机械臂路径优化等，提高检测物体的稳定性，增加数据集物体种类，提高抓取速度，将该技术推广应用到实际生产中。

参考文献：

[1] 杜学丹，蔡莹皓，鲁涛，等. 一种基于深度学习的机械臂抓取方法[J]. 机器人, 2017, 39(6): 820-828. (Du Xuedan, Cai, Yinghao Lu Tao. A robotic grasping method based on deep learning [J]. Robot, 2017, 39 (6): 820-828.)

[2] Hosoda K, Asada M. Versatile visual servoing without knowledge of true jacobian [C]// Proc of IEEE/RSJ/GI International Conference on. Intelligent Robots and Systems. Berlin: Springer, 1994: 186-193.

[3] Su Jianbo, Zhang Yanjun, Luo Zhiwei. Online estimation of image Jacobian matrix for uncalibrated dynamic hand-eye coordination [J]. International Journal of Systems, Control and Communications, 2008, 1 (1): 31-52.

[4] Horaud R, Dornaika F, Espiau B. Visually guided object grasping [J]. IEEE Trans on Robotics and Automation, 1998, 14(4): 525-532.

[5] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systems. 2012: 1097-1105.

[6] Johns E, Leutenegger S, Davison A J. Deep learning a grasp function for grasping under gripper pose uncertainty [C]//Procof IEEE/RSJ International Conference on Intelligent Robots and Systems.2016: 4461-4468.

[7] Zeng A, Yu Kuanting, Song Shuran, *et al*. Multi-view self-supervised deep learning for 6d pose estimation in the amazon picking challenge [C]// IEEE International Conference on Robotics and Automation. Piscataway,NJ: IEEE Press, 2017: 1386-1383.

[8] Levine S, Pastor P, Krizhevsky A, *et al*. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection [J]. International Journal of Robotics Research, 2018, 37(4-5): 421-436.

[9] Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proc of IEEE Conference On Computer Vision and Pattern Recognition. Washington DC:: IEEE Computer Society, 2014: 580-587.

[10] Uijlings J R R, De Van Sande K E A, Gevers T, *et al*. Selective search for object recognition [J]. International journal of computer vision, 2013, 104 (2): 154-171.

[11] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al*. Deep residual learning for image recognition [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC:: IEEE Computer Society, 2016: 770-778.

[12] Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC:: IEEE Computer Society, 2016: 779-788.

[13] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC:: IEEE Computer Society, 2017: 6517-6525.

[14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL].(2015-04-10). https://arxiv.org/abs/1409.1556.

[15] Lin Min, Chen Qiang, Yan Shuicheng. Network in network [EB/OL]. (2013-12-16) [2018-10-22]. https://arxiv. org/abs/1312. 4400

[16] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [C]//Proc of International Conference on Machine Learning. [S. l.] :CRC Press, 2015.

[17] Darknet: open source neural networks in C[EB/OL]. [2018-05-05]. https://pjreddie.com/darknet/.

[18] Arthur D, Vassilvitskii S. K-means+: the advantages of careful seeding [C]// Proc of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms.Philadelphia, PA: Society for Industrial and Applied Mathematics, 2007: 1027-1035.